中央大学 (2024/06/08)

No.1

講義内容

- 3.3 Gauss確率変数列
  - 3.3.1 non Gauss型確率変数列モデル
  - 3.3.2 スパース適応的しきい値推定量

Gauss確率変数列モデル

$$Y_j = \beta_j + \varepsilon_j \quad (j = 1, 2, \ldots, n) \quad (3.9)$$

ただし、$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2) \quad (\sigma > 0)$

β_jは未知の非確率的定数

σは既知or未知

3.3.1 劣Gauss型確率変数列モデル

固定デザインの線型回帰モデル

$$\underline{Y} = \underline{X}\,\underline{\beta}^* + \underline{\varepsilon}$$

ただし $\underline{Y} = (Y_1, \ldots, Y_n)^T$, $\underline{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$

$\underline{X}$: $n \times d$ のデザイン行列 $\underline{\beta}^* = (\beta_1^*, \ldots, \beta_d^*)^T$

$$\underline{X} = \begin{pmatrix} \underline{x}_1^T \\ \vdots \\ \underline{x}_n^T \end{pmatrix};\quad \underline{x}_j \in \mathbb{R}^d$$

ここで $n = d$

$$\underline{x}_j = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ 0 \end{pmatrix} \leftarrow j \implies (3.9)$$

ORT仮定

$$\frac{\underline{X}^T \underline{X}}{n} = \underline{I}_d = \overbrace{\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}}^{d}$$

ORT仮定 $\Longrightarrow$ $d \leq n$ （したがって、$d > n$ はダメ）

$\underline{X} = (\underline{x}_{1*}, \underline{x}_{2*}, \ldots\ldots, \underline{x}_{d*})$; $x_{j*} \in \mathbb{R}^n$
$(j = 1, 2, \ldots, n)$

と書いたとき

ORT $\Longrightarrow$ $\underline{x}_{i*}^T x_{j*} = \begin{cases} n & (i = j) \\ 0 & (i \neq j) \end{cases}$.

No.5

$\underline{Y} = \underline{X}\beta^* + \underline{\varepsilon}$ を次のように書きかえる

$$\underset{d\times 1}{\underline{y}} := \frac{1}{n}\underset{d\times n}{X^T}\, Y = \frac{1}{n}X^T(\underline{X}\beta^* + \underline{\varepsilon}) = \beta^* + \underline{\xi}$$

ただし. $\underline{\xi} = (\xi_1, \ldots, \xi_d)^T \sim \mathrm{subG}_d\left(\frac{\sigma^2}{n}\right)$.

ORT のもと回帰モデルは (3.9) に帰着.

No. 6

$\beta^* \in \mathbb{R}^d$ に対して

$$\|y - \beta^*\|_{2,d}^2 = \left\|\frac{1}{n} X^T Y - \beta^*\right\|_{2,d}^2$$

$$= \underbrace{\beta^{*T}\beta^*}_{\frac{1}{n}X^TX = I_d} - \frac{2}{n}\beta^{*T}X^TY + \underbrace{\frac{1}{n^2}Y^TXX^TY}_{\frac{1}{n}XX^T =}$$

$$= \frac{1}{n}\|X\beta^*\|_{2,d}^2 - \frac{2}{n}(X\beta^*)^T Y$$

$$+ \frac{1}{n}\|Y\|_{2,n}^2 \quad =: Q$$

$$+ \boxed{\frac{1}{n^2}Y^TXX^TY - \frac{1}{n}\|Y\|_{2,n}^2}$$

$$= \frac{1}{n} | \underline{Y} - \underline{X}\beta^* |_{2,n}^2 + Q$$

ただし,

$$Q = \frac{1}{n^2} \underline{Y}^T \underline{X}\,\underline{X}^T \underline{Y} - \frac{1}{n} |\underline{Y}|_{2,n}^2 .$$

よって, 最小2乗推定量 (LSE) $\hat{\beta}^{LS}$ は $\underline{y}$

次のモデルを考える

$$\underline{y} = \beta^* + \underline{\xi} \; ; \quad \underline{\xi} \sim \mathrm{sub}G_d\left(\frac{\sigma^2}{n}\right). \quad (3.10)$$

3.3.2 スパース適応的しきい値推定量

$\beta^* \in \mathbb{R}^d$ は k スパース $(1 \le k \le d)$

β^* の要素の $(d-k)$ 個はゼロ

すると、系 3.12 より、$^\forall \delta > 0$ に対して

$$\Pr\left(\mathrm{MSE}(\underline{X}\hat{\beta}^{LS}_{B_0(k)}) \le C_\delta \frac{\sigma^2 k}{n} \log\left(\frac{ed}{2k}\right)\right) \ge 1-\delta.$$

ただし、$\hat{\beta}^{LS}_{B_0(k)}$ は $B_0(k) = \{\beta \in \mathbb{R}^d ; |\beta|_{0,d} \le k\}$

のもとでの LSE

k を知っている。← 反則

No.9

一方, $\underline{y} = \beta^* + \underline{\xi}$ のもと, $\underbrace{\beta^* \text{を} \underline{y} \text{で推定する}}_{\text{事前情報がないとき}}$ のが普通 ex. 正規化

すると, $\forall \delta > 0$ に対して.

$$\Pr\left(\mathrm{MSE}(\underline{X}\hat{\beta}^{LS}) = |\underline{y} - \beta^*|_{2,d}^2 = |\underline{\xi}|_{2,d}^2 \leq C_\delta \frac{\sigma^2 d}{n} \right)$$

$$\geq 1 - \delta$$

No.10

$k = cd$ $(c \leq 1)$ のとき，k は d より十分小さいときを考える．

$j \in \{1, 2, \ldots, d\}$ に対して

$$\beta_j^* = 0 \qquad \left(\beta^* = (\beta_1^*, \ldots, \beta_d^*)^T \right)$$

のとき，$y_j = \xi_j$ である． $\underline{y} = (y_1, \ldots, y_d)^T$

すると，$y_j \sim \mathrm{subG}(\sigma^2/n)$ $\underline{\xi} = (\xi_1, \ldots, \xi_d)^T$

なので，補題 2.10 から，$^\forall t > 0$ に対して

$$\Pr\left(|\xi_j| \geq t\right) \leq 2\exp\left(-\frac{nt^2}{2\sigma^2}\right)$$

ここで

$$\delta := 2\exp\left(-\frac{t^2}{2\sigma^2}\right) \Leftrightarrow t = \sigma\sqrt{\frac{2\log\left(\frac{2}{\delta}\right)}{n}}$$

よって、$\forall \delta > 0$ に対して

$$\Pr\left(|\xi_j| \leq \sigma\sqrt{\frac{2\log\left(\frac{2}{\delta}\right)}{n}}\right) \geq 1-\delta. \quad (3.11)$$

$|\xi_j|$ は $\sigma\sqrt{\frac{2\log\left(\frac{2}{\delta}\right)}{n}}$ を超える確率は低い。

よって、ある $\tau > 0$ に対して

$$|y_j| \underbrace{\gg}_{\text{かなり大き、}} \tau \implies \hat{\beta}_j^* \neq 0 \text{ となる.}$$

またたとえば

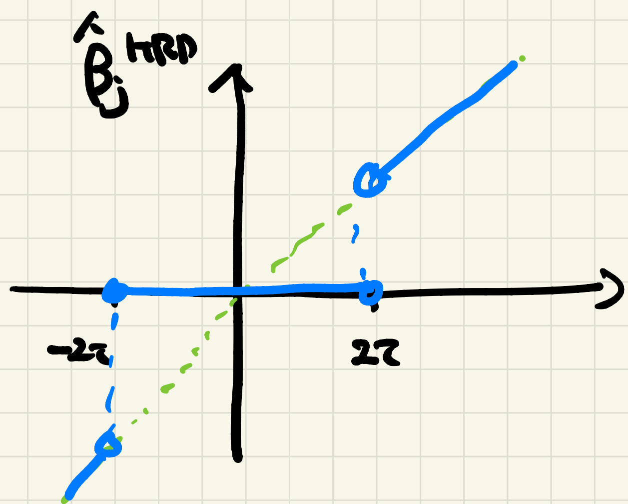
$$|y_j| \leq 2\tau \implies \beta_j^* = 0 \text{ となる.}$$

← 後で定める.

定義 3.13

$$\hat{\beta}_j^{HRD} = \begin{cases} y_j & (|y_j| > 2\tau) \\ 0 & (|y_j| \leq 2\tau) \end{cases}$$

$$= y_j \mathbb{I}\{|y_j| > 2\tau\}$$



τを決めるために、$\max\limits_{1\le j\le d} |y_j| \le \tau$ を評価

命題2.24から

$$\Pr\left(\max_{1\le j\le d} |\xi_j| \ge t\right) \le 2\,d \exp\left(-\frac{n t^2}{2\sigma^2}\right)$$

$$\delta := 2d \exp\left(-\frac{nt^2}{2\sigma^2}\right) \iff t = \sigma \sqrt{\frac{2\log\left(\frac{2d}{\delta}\right)}{n}}$$

よって、${}^\forall \delta > 0$に対して

$$\Pr\left(\max_{1\le j\le d} |\xi_j| \le \sigma \sqrt{\frac{2\log(2d/\delta)}{n}}\right) \ge 1-\delta.$$

定理 3.14 $\underline{Y} = \underline{X}\underline{\beta}^* + \underline{\varepsilon}$ は ORT 条件 $\underline{X}^T\underline{X} = n\underline{I}_d$

をみたすと仮定する。ただし、$\underline{\varepsilon} \sim \mathrm{subG}_n(\sigma^2)$ である。

$\underline{y} = \underline{X}^T\underline{Y}$ としたとき、$\underline{y} = \underline{\beta}^* + \underline{\xi}$; $\underline{\xi} \sim \mathrm{subG}_d(\sigma^2/n)$

である。$^\forall \delta > 0$ に対して

$$2\tau = 2\sigma\sqrt{\frac{2\log(2d/\delta)}{n}}$$

とし

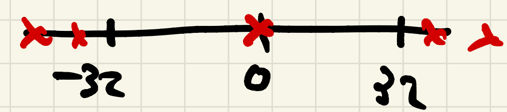
$$\hat{\beta}^{HRD}_j = \begin{cases} y_j & (|y_j| \geq 2\tau) \\ 0 & (|y_j| < 2\tau) \end{cases}$$

は以下の性質をみたす。

(1) $|\beta^*|_{0,d} = k$ とき

$$\Pr\left(\mathrm{MSE}(\underline{X}\hat{\beta}^{HRD}) = \frac{1}{n}|\underline{X}\hat{\beta}^{HRD} - \underline{X}\beta^*|_{2,d}^2 \lesssim \sigma^2 \frac{k}{n}\log\left(\frac{2d}{\delta}\right)\right) \geq 1-\delta.$$

(2) $\underbrace{\min_{j \in \mathrm{supp}(\underline{\beta}^*)}}$ $|\beta_j| > 3\tau$ のとき



← ゼロでないもの絶対値

のとき

$$\Pr\left(\underbrace{\mathrm{supp}(\hat{\beta}^{HRD}) = \mathrm{supp}(\beta^*)} \right) \geq 1-\delta$$

ゼロの場所を当てている

証明 (1)を示す.

$$A=\left\{\max_{1\le j\le d}|\xi_j|\le\tau\right\}=\left\{\omega\in\Omega;\ \max_{1\le j\le d}|\xi_j(\omega)|\le\tau\right\}.$$

命題2.24より

$$\Pr(A)\ge 1-\delta$$

また、$\underbrace{A\text{上}}_{\max|\xi_j|\le\tau}$で $j\in\{1,2,\dots,d\}$に対して、以下が成立する

$$|y_j|>2\tau \implies |\beta_j^*|=|y_j-\xi_j|\ge \overset{>2\tau}{\boxed{|y_j|}}\ \overset{\ge -\tau}{\boxed{-|\xi_j|}}>\tau \qquad \text{(a)}$$

$$|y_j|\le 2\tau \implies |\beta_j^*|=|y_j-\xi_j|\le|y_j|+|\xi_j|\le 3\tau \qquad \text{(b)}$$

No.18

よって, $j \in \{1, 2, \ldots, d\}$ に対して, A 上で

$$|\hat{\beta}_j^{\mathrm{HARD}} - \beta_j^*| = |\hat{\beta}_j^{\mathrm{HARD}} - \beta_j^*| \mathbb{I}\{|y_j| > 2\tau\} + |\hat{\beta}_j^{\mathrm{HARD}} - \beta_j^*| \mathbb{I}\{|y_j| \leq 2\tau\}$$

$$= |y_j - \beta_j^*| \mathbb{I}\{|y_j| > 2\tau\} + |\beta_j^*| \mathbb{I}\{|y_j| \leq 2\tau\}$$

$\because)$ $\hat{\beta}_j^{\mathrm{HARD}}$ の定義

$$\leq \underset{\because\ (a)}{|\xi_j| \mathbb{I}\{|\beta_j^*| > \tau\}} + \underset{\because\ (b)}{|\beta_j^*| \mathbb{I}\{|\beta_j^*| \leq 3\tau\}}$$

$$\leq \tau \mathbb{I}\{|\beta_j^*| > \tau\} + |\beta_j^*| \mathbb{I}\{|\beta_j^*| \leq 3\tau\}$$

(A 上で $|\xi_j| \leq \tau$)

$$\le 4 \min\{|\beta_j^*|, \tau\}$$

この式より

$$\|\hat{\beta}^{HRD} - \beta^*\|_{2,d}^2 = \sum_{j=1}^{d} |\hat{\beta}_j^{HRD} - \beta_j^*|^2$$

$$\le 16 \sum_{j=1}^{n} \min\{|\beta_j^*|^2, \tau^2\}$$

$$\le 16 \underbrace{\|\beta^*\|_{0,d}}_{=k} \boxed{\tau^2} = 2\sigma^2 \frac{\log\left(\frac{2d}{\delta}\right)}{n}$$

$$= \frac{32 k \sigma^2}{n} \log\left(\frac{2d}{\delta}\right) \lesssim \frac{\sigma^2 k \log\left(\frac{2d}{\delta}\right)}{n}$$

よって、 A が起これば、

$$\mathrm{MSE}(\underline{X}\hat{\underline{\beta}}^{HRD}) = |\hat{\underline{\beta}}^{HRD} - \underline{\beta}^*|_{2,d}^2 \lesssim \frac{\sigma^2 k \log\left(\frac{2d}{\delta}\right)}{n}$$

$\rightsquigarrow$

$$A \subset \left\{ |\hat{\underline{\beta}}^{HRD} - \underline{\beta}^*|_{2,d}^2 \lesssim \frac{\sigma^2 k \log\left(\frac{2d}{\delta}\right)}{n} \right\}$$

$\rightsquigarrow$

$$\Pr\left(|\hat{\underline{\beta}}^{HRD} - \underline{\beta}^*|_{2,d}^2 \lesssim \frac{\sigma^2 k \log\left(\frac{2d}{\delta}\right)}{n} \right) \geq \Pr(A) \geq 1 - \delta.$$

$\mathrm{supp}(\underline{\beta}^*) = \mathrm{supp}(\hat{\underline{\beta}}^{HRD}) \Leftrightarrow$ $\mathrm{supp}(\underline{\beta}^*) \subset \mathrm{supp}(\hat{\underline{\beta}}^{HRD})$ ①

かつ

$\mathrm{supp}(\underline{\beta}^*) \supset \mathrm{supp}(\hat{\underline{\beta}}^{HRD})$ ②

(2) の証明 ① $\beta_j^* \neq 0$ ならば，仮定から

$|\beta_j^*| > 3\tau$ であることを注意する．

$\max|\xi_i| < \tau \rightsquigarrow -|\xi_j| \geq -\tau$

A上では

$$|y_j| = |\beta_j^* + \xi_j| \geq |\beta_j^*| - |\xi_j| > 3\tau - \tau = 2\tau.$$

よって，$\hat{\beta}_j^{HRD}$ の定義より

$$\hat{\beta}_j^{HRD} \neq 0. \rightsquigarrow \mathrm{supp}(\beta^*) \subset \mathrm{supp}(\hat{\beta}^{HRD}).$$

② $\hat{\beta}_j^{HRD} \neq 0$ のとき

$$|\hat{\beta}_j^{HRD}| = |y_j| > 2\tau.$$

A 上では

$$|\beta_j^*| = |y_j - \xi_j| \geq |y_j| - |\xi_j| \geq 2\tau - \tau = \tau.$$

よって，$\beta_j^* \neq 0$ なので

$$\mathrm{supp}(\beta^*) \supset \mathrm{supp}(\hat{\beta}^{HRD})$$

以上より、 $A \Rightarrow \mathrm{supp}(\beta^*) = \mathrm{supp}(\hat{\beta}^{HRD})$

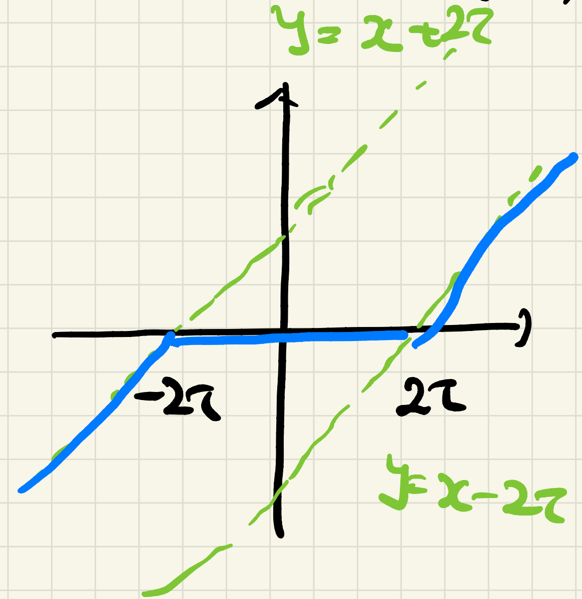
$\curvearrowright \quad A \subset \{ \mathrm{supp}(\hat{\beta}^{HRD}) = \mathrm{supp}(\beta^*) \}$

$$\Pr\left(\mathrm{supp}(\hat{\beta}^{HRD}) = \mathrm{supp}(\beta^*) \right) \geq \Pr(A) \geq 1 - \delta.$$

証

軟しきい値推定量

$$
\hat{\beta}_j^{SFT} = \begin{cases} y_j - 2\tau & (y_j > 2\tau) \\ y_j + 2\tau & (y_j < -2\tau) \\ 0 & (|y_j| \leq 2\tau) \end{cases}
$$



$$
\hat{\beta}_j^{SFT} = \left(1 - \frac{2\tau}{|y_j|}\right) y_j \; ; \quad a_+ = \max(a, 0)
$$

定理 3.15

(1) $|\beta^*|_{0,d} = k$ のとき, $^\forall \delta > 0$ に対して

$$\Pr\left(\mathrm{MSE}(\underline{X}\hat{\beta}^{SF7}) = \frac{1}{n}|\underline{X}\hat{\beta}^{SF7} - \underline{X}^*\beta^*|_{2,n}^2 \lesssim \frac{\sigma^2 k \log\left(\frac{2d}{\delta}\right)}{n}\right)$$

$$\geq 1-\delta.$$

(2) $\displaystyle\min_{1\leq j\leq d}|\beta_j^*| > 3\tau$ のとき, $^\forall \delta > 0$ に対して

$$\Pr\left(\mathrm{supp}(\hat{\beta}^{SF7}) = \mathrm{supp}(\beta^*)\right) \geq 1-\delta.$$

証明の概略

(1) 事象 $A = \{\max |S_j| \le \tau\}$ 上では

$$|\hat{\beta}_j^{SFT} - \beta_j^*| \le 5 \min\{|\beta_j^*|, \tau\} \quad (j = 1, \dots, d).$$

よって

$$\|\hat{\beta}^{SFT} - \beta^*\|_{2,d}^2 \le \frac{50 k \sigma^2 \log\left(\frac{2d}{\delta}\right)}{n}.$$

(2) A 上では

① $\beta_j^* \ne 0 \Longrightarrow |y_j \pm 2\tau| > 0 \Longrightarrow \hat{\beta}_j \ne 0.$

② $\hat{\beta}_j \ne 0 \Longrightarrow |\beta_j^*| \ge \tau \Longrightarrow \beta_j^* \ne 0.$ $\square$

以上のまとめ

$$\Pr\left(\frac{1}{n}\|\underline{X}\hat{\beta}^{LS}-\underline{X}\beta^*\|_{2,n}^2 \lesssim \frac{\sigma^2 d}{n}\right) \geq 1-\delta$$

$$\Pr\left(\frac{1}{n}\|\underline{X}\hat{\beta}^{LS}_{B_0(R)}-\underline{X}\beta^*\|_{2,n}^2 \lesssim \frac{\sigma^2 k}{n}\log\left(\frac{ed}{2k}\right)\right) \geq 1-\delta$$

$$\boxed{\frac{e}{2} < 2}$$

ORT条件のもと

$$\Pr\left(\frac{1}{n}\|\underline{X}\hat{\beta}^{HT}-\underline{X}\beta^*\|_{2,n}^2 \lesssim \frac{\sigma^2 k}{n}\log\left(\frac{2d}{\delta}\right)\right) \geq 1-\delta$$

$$\Pr\left(\frac{1}{n}\|\underline{X}\hat{\beta}^{SFT}-\underline{X}\beta^*\|_{2,n}^2 \lesssim \frac{\sigma^2 k}{n}\log\left(\frac{2d}{\delta}\right)\right) \geq 1-\delta$$

3.4 高次元線型回帰モデル

$\underset{\sim}{y} = \underset{\sim}{\beta}^* + \underset{\sim}{\xi}$ のとき

$$\hat{\beta}_j^{HRD} = y_j \mathbb{I}\{|y_j| > 2\tau\},\quad \underset{\sim}{\hat{\beta}}^{HRD} = (\hat{\beta}_1^{HRD}, \cdots, \hat{\beta}_d^{HRD})^T$$

は関数

$$\underset{\sim}{\beta} \longmapsto |\underset{\sim}{y} - \beta|_{2,d}^2 + 4\tau^2 |\underset{\sim}{\beta}|_{0,d}$$

を最小にする点である。

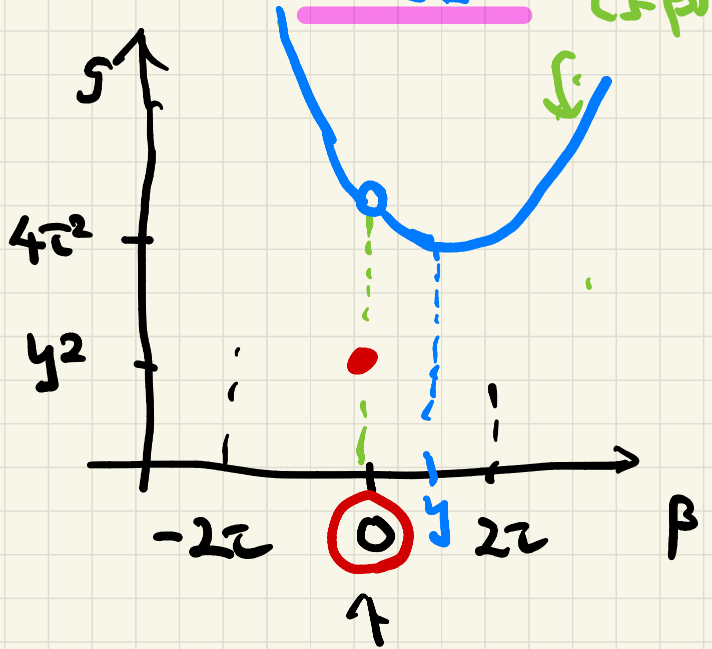
簡単のため $d=1$ とする.

$$g(\beta)=(y-\beta)^2+4\tau^2\mathbb{1}\{\beta\neq 0\}$$

すなわち

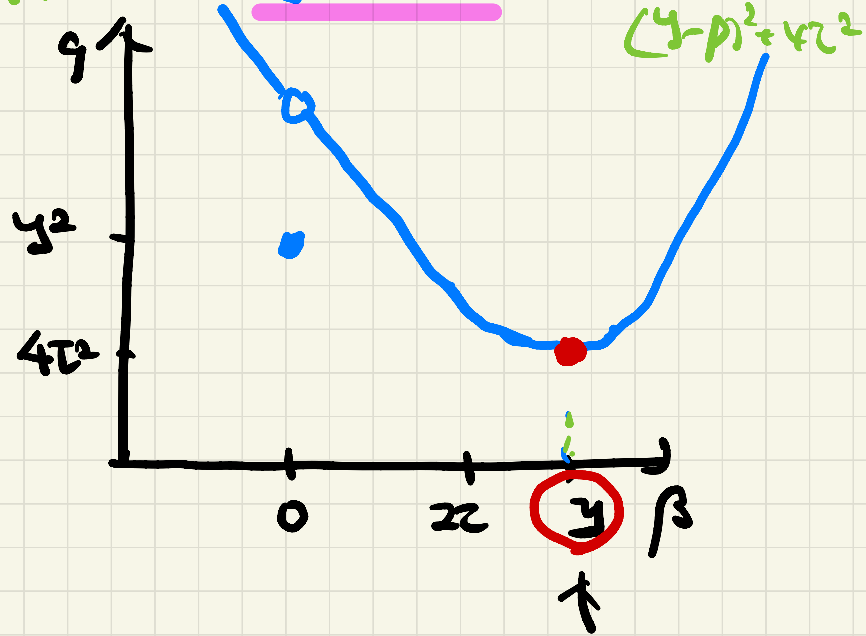
$$g(\beta)=\begin{cases}(y-\beta)^2+4\tau^2 & (\beta\neq 0)\\ y^2 & (\beta=0)\end{cases}$$

$0<y\leq 2\tau$



最小となる点

$y>2\tau$



最小となる点

関数 $\beta \mapsto g(\beta)$ は不連続関数かつ凸でない

$$\hat{\beta}_j^{\mathrm{SFT}} = \left(1 - \frac{2\tau}{|y_j|}\right)_+ y_j, \quad \hat{\underline{\beta}}^{\mathrm{SFT}} = \left(\hat{\beta}_1^{\mathrm{SFT}}, \ldots, \hat{\beta}_d^{\mathrm{SFT}}\right)^T$$ No.30

は関数

$$\underline{\beta} \mapsto |\underline{y} - \beta|_{2,d}^2 + 4\tau|\underline{\beta}|_{1,d}$$

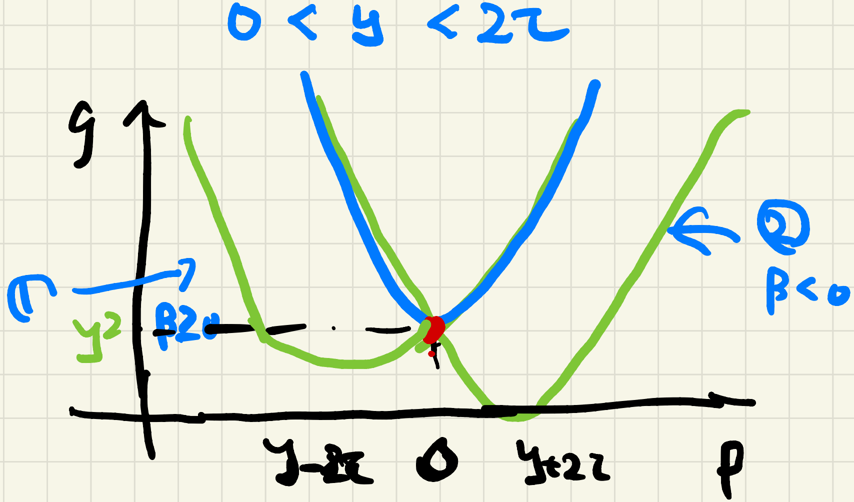
を最小にする点である.

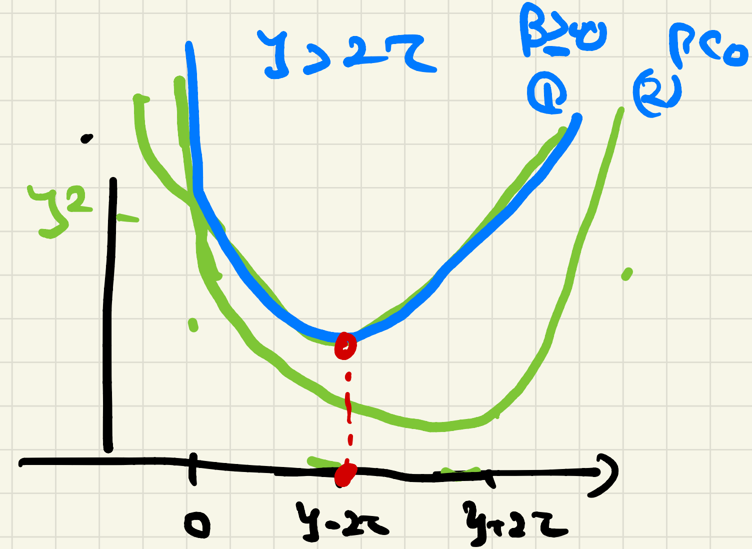
$d=1$の場合

$$g(\beta) = (y-\beta)^2 + 4\tau|\beta|$$

$$= \begin{cases} (y-\beta)^2 + 4\tau\beta & (\beta \geq 0) \\ (y-\beta)^2 - 4\tau\beta & (\beta < 0) \end{cases}$$

$$= \begin{cases} \{\beta-(y-2\tau)\}^2 - (y-2\tau)^2 \\ \{\beta-(y+2\tau)\}^2 - (y+2\tau)^2 \end{cases}$$ ①





関数 $\beta \mapsto g(\beta)$ は 凸 である　　No.32

かつ 連続

凸関数の最適化は容易！

ORT条件のもとでは

$\hat{\beta}^{HRD}$ は関数

$$\underline{\beta} \longmapsto |\underline{Y} - \underline{X}\beta|_{2,n}^{2} + 4\tau^{2}|\beta|_{0,d}$$

を最小化する $\underline{\beta}$

$\hat{\beta}^{SFD}$ は関数

$$\underline{\beta} \longmapsto |\underline{Y} - \underline{X}\beta|_{2,n}^{2} + 4\tau|\beta|_{1,d}$$

を最小化する $\underline{\beta}$

$\hat{\beta}^{HRD}$ は $\beta \longmapsto \underbrace{(\text{残差平方和}) + (\text{非凸な罰則})}_{\text{連続でない}}$

$\hat{\beta}^{SFT}$ は $\beta \longmapsto \underbrace{(\text{残差平方和}) + (\text{凸な罰則})}_{\text{連続}}$

ORT 条件をみたさないとき

$\underline{y} = \beta^* + \underline{\xi}$ に帰着できない!

$\hat{\beta}^{BIC}$ は関数

$$\beta \longmapsto \frac{1}{n}|\underline{y} - \underline{X}\beta|_{2,n}^2 + \tau^2|\beta|_{0,d}$$

を最小にする $\underline{\beta}$

$\hat{\beta}^{\alpha}$ は関数

$$\beta \longmapsto \frac{1}{n}|\underline{y} - \underline{X}\beta|_{2,n}^2 + 2\tau|\beta|_{1,d}$$

を最小にする $\underline{\beta}$

注意3.18（計算コスト）

(1) $\hat{\beta}^{BIC}$ は計算量がぼう大 → brutal

$$\min_{\beta \in \mathbb{R}^d} \left\{ \frac{1}{n} |Y - X\beta|_{2,n}^2 + \tau^2 |\beta|_{0,d} \right\}$$

$$= \min_{1 \le k \le d} \left\{ \min_{\beta \in \mathbb{R}^d : |\beta|_{0,d} = k} \left(\quad \cdot \quad + \tau^2 k \right) \right\}$$

$\displaystyle \min_{\beta \in \mathbb{R}^d : |\beta|_{0,d} = k} |Y - X\beta|_{2,n}^2$ は $\binom{d}{k}$ 回の計算

1つのLSEの計算に $O(k^3)$ 回

よって、計算コストは

$$C\sum_{k=0}^{d}\binom{d}{k}k^3 = C d^3 2^d$$

(2) $\hat{\beta}^L$ の計算コスト

効率的なものが知られている

(1) 座標降下法 → R package glmnet

(2) LARS (least angle regression)

(3) FISTA algorithm

(4) 交互方向降下法

川野他 (2018, pp. 12–20)「スパース推定法による統計モデリング(共立)」を参照.

3.4.2 $\hat{\beta}^{BIC}$ の精度評価（ベンチマーク）

ORT条件は必要なし!

命題3.19 $\underline{Y} = \underline{X}\beta^* + \underline{\varepsilon}$, $\underline{\varepsilon} \sim \mathrm{subG}_n(\sigma^2)$ として,

$|\beta^*|_{0,d} \geq 1$ とする. このとき, 罰則

$$\tau^2 = 16\log(12)\frac{\sigma^2}{n} + 32\sigma^2\frac{\log(ed)}{n}$$

と $^\forall \delta > 0$ に対し

$$\Pr\left(\frac{1}{n}|\underline{X}\hat{\beta}^{BIC} - \underline{X}\beta^*|_{2,n}^2 \lesssim |\beta^*|_{0,d}\,\sigma^2\frac{\log\left(\frac{ed}{\delta}\right)}{n}\right) \geq 1-\delta.$$

証明の概略　$\hat{\beta}^{BIC}$ の定め方より

$$\frac{1}{n}\|\underline{Y}-\underline{X}\hat{\underline{\beta}}^{BIC}\|_{2,n}^2+\tau^2\|\hat{\underline{\beta}}^{BIC}\|_{0,d}\leq\frac{1}{n}\|\underline{Y}-\underline{X}\underline{\beta}^*\|_{2,n}^2+\tau^2\|\underline{\beta}^*\|_{0,d}$$

である．代数的な計算から

$$\|\underline{X}\hat{\underline{\beta}}^{BIC}-\underline{X}\underline{\beta}^*\|_{2,n}^2\leq n\tau^2\|\underline{\beta}^*\|_{0,d}+\underbrace{2\underline{\varepsilon}^T X(\hat{\underline{\beta}}^{BIC}-\underline{\beta}^*)}_{=:A}-n\tau^2\|\hat{\underline{\beta}}^{BIC}\|_{0,d} \quad (※)$$

すると

$$A=\underbrace{\boxed{2\underline{\varepsilon}^T\left(\frac{X(\hat{\underline{\beta}}^{BIC}-\underline{\beta}^*)}{\|\underline{X}(\hat{\underline{\beta}}^{BIC}-\underline{\beta}^*)\|_{2,n}}\right)}}_{=a}\underbrace{\boxed{\|\underline{X}\hat{\underline{\beta}}^{BIC}-\underline{X}\underline{\beta}^*\|_{2,n}}}_{=b}$$

$$\leq 2\left[\underline{\varepsilon}^T\left(\frac{\underline{X}\hat{\beta}^{BIC}-\underline{X}\beta^*}{|\qquad|_{2,n}}\right)\right]^2+\frac{1}{2}|\qquad|_{2,n}^2$$

↪ (*)をAに代入

$\because$) $2ab \leq 2a^2+\frac{1}{2}b^2$

ここで, $\underline{z}\in\mathbb{R}^d$ に対して

$$u(\underline{z})=\frac{\underline{z}}{|\underline{z}|_{2,d}}$$

とおくと

$$|\underline{X}\hat{\beta}^{BIC}-\underline{X}\beta^*|_{2,n}^2 \leq 2n\tau^2|\beta^*|_{0,d}+4\left[\underline{\varepsilon}^T u(\underline{X}\hat{\beta}^{BIC}-\underline{X}\beta^*)\right]^2$$

$$-2n\tau^2|\hat{\beta}^{BIC}|_{0,d} \qquad (3.16)$$

sup-out

$$4\left[\underline{\varepsilon}^T U(\underline{X}\hat{\beta}^{BIC} - \underline{X}\beta^*)\right]^2 - 2n\tau^2 |\hat{\beta}^{BIC}|_{0,d}$$

$$\leq \max_{1\leq k\leq d} \left\{ \max_{S\subset\mathbb{R}^d : \#(S)=k} \sup_{\#(S)=k} \sup_{\underline{u}\in B_2^{r_{S+}}} \left[4\left[\underline{\varepsilon}^T \Psi_{S+} \underline{u}\right]^2 - 2n\tau^2 k \right] \right\}$$

ただし、$\underline{c}_j \in \mathbb{R}^n$ を $\underline{X}$ の j 列 $(j=1,\ldots,d)$ としたとき

$$\Psi_{S+} = [\psi_1, \ldots, \psi_{r_{S+}}] ; \quad \psi_j \in \mathbb{R}^n \ (j=1,\ldots,S+)$$

は $\mathrm{span}\{c_j : j\in S \cup \mathrm{supp}(\beta^*)\}$ の正規直交基底，$r_{S+} \leq \#(S) + |\beta^*|_{0,d}$

$$\Pr\left(\qquad \geq t \right)$$

$$\leq \sum_{k=1}^{d} \sum_{S \subset \mathbb{R}^d : \#(S)=k} \Pr\left(\sup_{\underline{u} \in B_2^{r_{St}}} \left[\underline{\xi}^T \Phi_{r_{St}} \underline{u} \right]^2 \geq \frac{t}{4} + \frac{t}{2} n\tau^2 k \right)$$

$$\leq \sum_{k=1}^{d} \sum_{S \subset \mathbb{R}^d : \#(S)=k} \exp\left(-\frac{t}{32\sigma^2} - 2k \log(ed) + |\beta^*|_{0,d} \log(12) \right)$$

∵) 命題2.30のε網と最大不等式

$$\leq \sum_{k=1}^{d} \left(\frac{1}{ed} \right) \exp\left(-\frac{t}{32\sigma^2} + |\beta^*|_{0,d} \log(12) \right)$$

よって

タイポ

$$\Pr\left(\underbrace{|\underline{X}\hat{\beta}^{BIC}-\underline{X}\beta^*|_{2,n}^2 \geq 2n\tau^2|\beta^*|_{0,d}+t}_{(3.16)\text{の左辺と右辺の3項目}}\right)$$

$$\leq \Pr\left(\underbrace{2n\tau^2|\beta^*|_{0,d}+4\left[\underline{\varepsilon}^T u(\underline{X}\hat{\beta}^{BIC}-\underline{X}\beta^*)\right]^2-2n\tau^2|\hat{\beta}^{BIC}|_{0,d}}_{|\underline{X}\hat{\beta}^{BIC}-\underline{X}\beta^*|_{2,n}^2 \leq \text{ by (3.16)}} \geq 2n\tau^2|\beta^*|_{0,d}+t\right)$$

$$\leq \underbrace{\sum_{k=1}^{d}\left(\frac{1}{ed}\right)^{k}}_{\leq 1}\underbrace{\exp\left(-\frac{t}{32\sigma^2}+|\beta^*|_{0,d}\log(12)\right)}_{=:\delta\text{とおいて、解き直す}}$$

□

証明の流れ

① $\hat{\beta}^{BIC}$ の定義を用いて

$$\|\underline{X}\hat{\beta}^{BIC} - \underline{X}\beta^*\|_{2,n}^2 \leq n\tau^2\|\beta^*\|_{0,d}$$

$$+ \underline{2\underline{\varepsilon}^T\underline{X}(\hat{\beta}^{BIC} - \beta^*) - n\tau^2\|\hat{\beta}^{BIC}\|}_{0,d}$$

sup-out

$\Downarrow$

最大不等式を使い評価

② $A = \{ \qquad \ge t \}$ としたとき

最大不等式を用いると

$$\Pr(A) \le \underbrace{\exp\left(-\frac{t}{32\sigma^2} + |\beta^*|_{0,d}\log(12)\right)}_{=\,\delta}$$

$$\Pr(\qquad < t) \ge 1-\delta$$

③ $A^c \Longrightarrow |\underline{X}\hat{\beta}^{BIC} - \underline{X}\beta^*|_{2,n}^2 \le 2n\tau^2 + t \lesssim |\beta^*|_{0,d}\sigma^2 \log\left(\frac{ed}{\delta}\right)$

より

$$1-\delta \le \Pr(A^c) \le \Pr\left(|\underline{X}\hat{\beta}^{BIC} - \underline{X}\beta^*|_{2,n}^2 \lesssim \qquad \right)$$

3.4.3 Lasso 推定量の精度評価

モデル $\underset{n\times 1}{\underline{Y}} = \underset{n\times d}{\underline{X}}\,\underset{d\times 1}{\underline{\beta}^*} + \underset{n\times 1}{\underline{\varepsilon}}$; $\underline{\varepsilon} \sim \mathrm{subG}_n(\sigma^2)$ $(\sigma > 0)$

を仮定する。さらに

$$\underline{X} = (\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_d) ; \quad X_j \in \mathbb{R}^n \ (j = 1, \ldots, n)$$

に対して

$$\max_{1\leq j\leq n} |\underline{X}_j|_{2,n} \leq \sqrt{n}$$

を仮定す。

$$|\underline{a}|_{2,n} = \sqrt{\sum_{j=1}^{n} a_j^2}$$

$$\underline{a} = (a_1, \ldots, a_n)^T$$

命題 3.20

$\forall \delta > 0$ に対して

$$2\tau := 2\sigma\sqrt{\frac{2\log(2d)}{n}} + 2\sigma\sqrt{\frac{2\log\left(\frac{1}{\delta}\right)}{n}}$$

として Lasso の推定量

$$\underline{\hat{\beta}}^{L} \in \arg\min_{\beta \in \mathbb{R}^d} \left\{ \|\underline{Y} - \underline{X}\beta\|_{2,n}^2 + 2\tau\|\beta\|_{1,d} \right\}$$

$$\|\beta\|_{1,d} = \sum_{j=1}^{d} |\beta_j| \; ; \; \underline{\beta} = (\beta_1, \ldots, \beta_d)^T$$

について

No.49

$$\Pr\left(\frac{1}{n}\|\underline{X}\hat{\beta}^{r}-\underline{X}\beta^{*}\|_{2,n}^{2}\leq 4\|\beta^{*}\|_{1,d}\,\sigma\sqrt{\frac{2\log(2d)}{n}}+4\|\beta^{*}\|_{1,d}\sqrt{\frac{\log\left(\frac{1}{\delta}\right)}{n}}\right)\geq 1-\delta$$

$$\left(\Pr\left(\frac{1}{n}\|\underline{X}\hat{\beta}^{r}-\underline{X}\beta^{*}\|_{2,n}^{2}\geq 4\|\beta^{*}\|_{1,2}\tau\right)\geq 1-\delta\right)$$

が成り立つ

証明　$\hat{\beta}^L$の定め方から　　　　　　　　　　　　　　　No.56

$$\frac{1}{n}\|\underline{Y}-\underline{X}\hat{\beta}^L\|_{2,n}^2+2\tau\|\hat{\beta}^L\|_{1,d}\leq\frac{1}{n}\|\underline{Y}-\underline{X}\beta^*\|_{2,n}^2+2\tau\|\beta^*\|_{1,d}\quad(a)$$

が成立。一方，(a)とHölderの不等式より

$$\|\underline{X}\hat{\beta}^L-\underline{X}\beta^*\|_{2,n}^2=\|\underline{X}\hat{\beta}^L-\underline{Y}+\underline{Y}-\underline{X}\beta^*\|_{2,n}^2$$

$$=\|\underline{Y}-\underline{X}\hat{\beta}^L\|_{2,n}^2+\|\underbrace{\underline{Y}-\underline{X}\beta^*}_{=\underline{\varepsilon}}\|_{2,n}^2$$

$$+2(\underbrace{\underline{Y}-\underline{X}\beta^*}_{=\underline{\varepsilon}})^T(\underline{X}\hat{\beta}^L\underbrace{-\underline{X}\beta^*-\varepsilon}_{=-\underline{Y}})$$

No.51

$$\leq -2n\tau|\hat{\beta}^{c}|_{1,d} + |\underline{Y} - \underline{X}\beta^*|_{2,n}^2 + 2n\tau|\beta^*|_{1,d}$$

$$= |\underline{\varepsilon}|_{2,n}^2$$

$$+ \cancel{|\underline{\varepsilon}|_{2,n}^2} + 2\underline{\varepsilon}^T(\underline{X}\hat{\beta}^{c} - \underline{X}\beta^*) - \cancel{2|\underline{\varepsilon}|_{2,n}^2}$$

$$= 2\underline{\varepsilon}^T(\underline{X}\hat{\beta}^{c} - \underline{X}\beta^*) + 2\tau n\{|\beta^*|_{1,d} - |\hat{\beta}^{c}|_{1,d}\}$$

$$= 2(\underline{X}^T\underline{\varepsilon})^T(\hat{\beta}^{c} - \beta^*) + 2\tau n\{|\beta^*|_{1,d} - |\hat{\beta}^{c}|_{1,d}\}$$

$$\leq 2|\underline{X}^T\underline{\varepsilon}|_{\infty}\{|\hat{\beta}^{c}|_{1,d} + |\beta^*|_{1,d}\}$$

Hölderの不等式 $$+ 2\tau n\{|\beta^*|_{1,d} - |\hat{\beta}|_{1,d}\}$$

$$= 2\{|\underline{X}^T\underline{\varepsilon}|_\infty - n\tau\}|\hat{\beta}^c|_{1,d}$$

$$+ 2\{|\underline{X}^T\underline{\varepsilon}|_\infty + n\tau\}|\beta^*|_{1,d}$$

である。ただし、$|\underline{a}|_\infty = \max\{|a|_1, \ldots, |a|_d\}$

$\underline{a} = (a_1, \ldots, a_d)^T$ である。

ここで、$j = 1, 2, \ldots, d$ に対し

$$|\underline{X}_j|_{2,n} \leq \sqrt{n} \quad ; \quad \underline{X} = (\underline{X}_1, \ldots, \underline{X}_d)$$

なので

$$\boxed{\underline{X}_j^T \underline{\varepsilon}} \sim \mathrm{subG}(\boxed{n\sigma^2})$$

となる.

∵) $s>0$ に対して

$$E\left[\exp\left(s\,\underline{X}_j^T \underline{\varepsilon}\right)\right] = E\left[\exp\left(s\,|\underline{X}_j|_{2,n} \frac{\underline{X}_j^T}{|\underline{X}_j|_{2,n}} \underline{\varepsilon}\right)\right]$$

$$\leq E\left[\exp\left(s\sqrt{n}\,\underline{u}^T \underline{\varepsilon}\right)\right]$$

$$\underline{u} = \frac{\underline{X}_j}{|\underline{X}_j|_{2,n}} \rightsquigarrow |u|_{2,n} = 1$$

$$\leq \exp\left(\frac{n}{2}\lambda^2\sigma^2\right) \quad (\because\ \varepsilon \sim \mathrm{subG}_n(\sigma^2))$$

$$\mathbb{E}[\exp(\lambda u^\top \varepsilon)]$$

$$\leq \exp\left(\frac{\lambda^2}{2}\sigma^2\right)$$

よって、命題2.24を思い出すと

No.56

$\boxed{w_j} \sim \mathrm{subG}(\boxed{\sigma^2})$ なので

$$\Pr\left(\max_{1\leq j\leq d} |w_j| > t\right) \leq 2d\exp\left(-\frac{t^2}{2\sigma^2}\right)$$

$\|X^T \varepsilon\|_\infty = \max_{1\le j\le n} |X_j^T \varepsilon^T|$ であるので

No.55

$$\Pr\left(\|X^T\varepsilon\|_\infty > t\right) \le 2d\exp\left(-\frac{t^2}{2n\sigma^2}\right) =: \delta$$

である。 ―― を解くと

$$t := \sigma\sqrt{2n\log(2d)} + \sigma\sqrt{2n\log\left(\frac{1}{\delta}\right)} = \tau_n$$

となる。さらに

$\|X^T\varepsilon\|_\infty \le t$ ならば

No.56

$$\|\underline{X}\hat{\beta}^{c} - \underline{X}\beta^{*}\|_{2,n}^{2} \leq 2\{\|\underline{X}^{T}\underline{\varepsilon}\|_{\infty} - n\tau\}\|\hat{\beta}\|_{1,d}$$
$$+ 2\{\|\underline{X}^{T}\underline{\varepsilon}\|_{\infty} + n\tau\}\|\beta^{*}\|_{1,d}$$

と合せると

$$\|\underline{X}\hat{\beta}^{c} - \underline{X}\beta^{*}\|_{2,n}^{2} \leq 2\underbrace{\{t - n\tau\}}_{=0}\|\hat{\beta}^{c}\|_{1,d}$$
$$+ 2\{\underbrace{t}_{=n\tau} + n\tau\}\|\beta^{*}\|_{1,d}$$
$$= 4n\tau\|\beta^{*}\|_{1,d}$$

以上から

$$|\underline{X}^T \underline{\varepsilon}| \leq t \Rightarrow |\underline{X}\hat{\beta}^{\lambda} - \underline{X}\beta^*|_{2,n}^2 \leq 4n\tau |\beta^*|_{1,d}$$

$$\{|\underline{X}^T \underline{\varepsilon}| \leq t\} \subset \{|\underline{X}\hat{\underline{\beta}}^{\lambda} - \underline{X}\underline{\beta}^*|_{2,n}^2 \leq 4n\tau|\beta^*|_{1,d}\}$$

$$\Pr(|\underline{X}^T \underline{\varepsilon}| \leq t) \leq \Pr(|\underline{X}\hat{\beta}^{\lambda} - \underline{X}\beta|_{2,n}^2 \leq 4n\tau|\beta^*|_{1,d})$$

No.57

$\Pr(|\underline{X}^T \underline{\varepsilon}| > t) \leq \delta$ であったことを

思い出せば

$$\Pr\left(|\underline{X}\hat{\underline{\beta}}^{\wedge} - \underline{X}\beta^*|_{2,n}^2 \leq 4\lambda\tau|\beta^*|_{1,d}\right) \geq 1-\delta$$

$\boxtimes$